

複数人による選択店舗の構造類似性に基づく飲食店推薦システム

谷口 雄大[†] 北山 大輔[†]

[†] 工学院大学 情報学部 コンピュータ科学科 〒163-8677 東京都新宿区西新宿 1-24-2

E-mail: tj113067@ns.kogakuin.ac.jp, kitayama@cc.kogakuin.ac.jp

あらまし 飲食店検索サイトで飲食店を決める際に、検索ワードとしてカテゴリや予算などを入力して飲食店を決定する。1人の場合では、自分の食べたい飲食店を選ぶことは容易だが2人以上のグループで飲食店を決める場合、予算やメンバーの好み異なるため、検索条件を決めることが難しい。そこで、本研究ではグループで飲食店を決める場合に、それぞれのユーザによる飲食店選択に基づき、飲食店を推薦する手法を提案する。まず、グループのメンバーそれぞれが選択した飲食店同士の類似度を算出する。その際に、「カテゴリ」「平均価格」「説明文」の3つの要素からなる店舗構造の類似度を定義した。類似度が最も高かった飲食店同士の情報から、グループとしての嗜好を抽出し、飲食店を推薦する。

キーワード 飲食店推薦, グループ推薦, 店舗構造の類似度

1. はじめに

ユーザが飲食店を利用する際に、同僚や家族、友人などグループで飲食店を訪れる機会が多い [1]。このような場合、飲食店検索サイト (HotPepper グルメ^(注1), ぐるなび^(注2), 食べログ^(注3)) 等で検索し、行きたい飲食店を決定することがある。しかし飲食店を決める際には様々な問題があげられる。グループの場合ではそれぞれの好みや、予算など様々な要素をうまくまとめる必要がある。このような問題を解決するために、個々人の店舗選択を用いて、グループの嗜好にあった飲食店を推薦する手法を提案する。

例えば、あるユーザ1は焼肉等の肉類を好んでいるが、イタリアンなどを好まないユーザだとする。もう一人のユーザ2は和食を好んでいるが、海鮮系を好まないユーザだったとする。このような際に、ユーザ1とユーザ2の飲食店リストから、お互いの予算や、好みではないカテゴリを考慮して、肉料理が特徴の和食の店舗などグループの嗜好に合う飲食店を推薦する。

本稿の構成を以下に示す。2章で関連研究について説明する。3章では想定されるデータについて説明する。4章では提案手法として、飲食店推薦システムについて説明する。5章では構造類似の実験について説明する。6章ではグループが好むペアに基づく推薦について説明する7章ではまとめと今後の課題を説明する。

2. 関連研究

グループ推薦を行う研究として、瀬占ら [2] は TV 番組情報から個々人の嗜好とグループでの行動履歴を用いて、グループ間のパワーバランスを推測したシステムからグループに適したコンテンツを推薦する手法を提案している。本研究では、店

舗という複数のメタデータを持つコンテンツを扱う。また伊藤ら [3] も EC サイトにおける商品の推薦として、グループの行動履歴を用いたコンテンツに基づいたフィルタリングを利用している。また、観光地推薦として奥蘭ら [4] はユーザが画像の選択をすることで嗜好を抽出している。複数嗜好の統合として各ユーザの嗜好を正規化しベクトルで表現している。

飲食店のジャンル選択として、神島 [5] は嗜好の予測として内容ベースフィルタリングを上げているが、これらは好みやジャンルを利用者に尋ねてその条件に合ったものを推薦するものである。ジャンルや好みは利用するが、好みを事前に明確にする方法は本研究とは異なっている。選択嗜好として、大坪 [6] はユーザの好みが明確でない状況を仮定し、飲食店を選択していく上で好みを明確にしていく手法を提案した。しかしこれらはグループに対しては考慮されていない。山田ら [7] はユーザに2択のメニューを提示し選択を繰り返すことでジャンルや嗜好を分析している。高玉ら [8] は商品を選択していく過程でユーザの好み具体的になり、また変化することから類似した別カテゴリの商品推薦システムを提案している。

レビュー類似度を扱っている研究として、山本らは [9] レビューから単語をクラスタリングし、レビューとの内積から類似度を計算している。レビューを利用しているが、本研究ではレビューではなく飲食店の紹介文を用いて単語の類似度を求めている。

3. 想定する店舗データ

本研究で扱う飲食店のデータは、カテゴリ、平均価格、説明文を持つ。カテゴリは、飲食店のジャンルを表すメタデータで、階層構造を持たず1店舗に複数ついてもよい。平均価格は、その店舗で使われる金額を表し、「2000円」など数値データを扱う。「1000~2000円」など、幅を持つ場合は平均値などで数値データに変換する。説明文は、店舗の特徴を表したテキストであり TF-IDF により文書ベクトルとして扱う。

TF-IDF の求め方として、まず飲食店の説明文を形態素解析

(注1) : <https://www.hotpepper.jp/>

(注2) : <https://www.gnavi.co.jp/>

(注3) : <https://tabelog.com/>

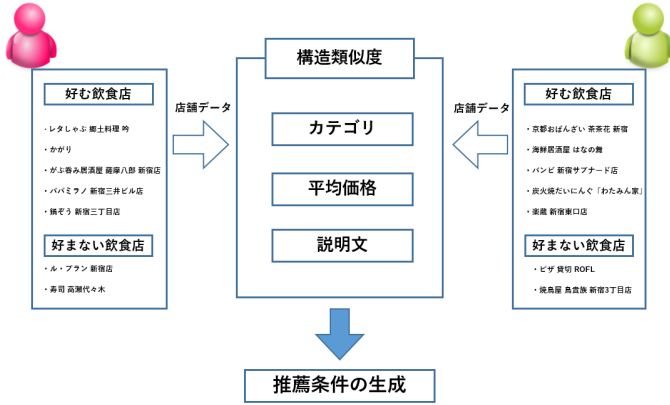


図1 システム全体像

する。取り出された単語集合を d として扱う，単語集合の中にある単語 n の出現回数を w_n^d として，以下の式で求める。

$$TF(n, d) = w_n^d \quad (1)$$

IDF はその店舗が属するカテゴリに限定して算出する。そのカテゴリのすべての店舗数を D とし，説明文に単語 n を含む店舗数を t_n として以下の式で求める。

$$IDF(n) = \log \frac{D}{t_n} \quad (2)$$

4. 提案手法

飲食店を推薦する際のシステムの流れを示す (図1)。

- (1) 各ユーザが飲食店を選択する。
- (2) ユーザが選んだ飲食店同士の類似度に基づき，グループの嗜好を示す飲食店を抽出する。
- (3) 抽出した店舗を用いて，推薦条件を決定し飲食店を推薦する。

ユーザが選択する飲食店は2種類ある。ユーザが好む飲食店とユーザが好まない(嫌いな)飲食店の2つであり，それぞれ複数選択できるものとする。

4.1 店舗の構造類似度

グループに適した飲食店を求めるために，本研究では選択された飲食店から3つの要素「カテゴリ」「平均価格」「説明文」を組み合わせることで，飲食店同士の類似度を定義する。

4.1.1 カテゴリの類似度

本研究ではカテゴリとして分類分けをされているが，カテゴリ同士でも似ているカテゴリ，似ていないカテゴリが存在するためカテゴリの類似度を予備実験により決定した。

本稿で使用するカテゴリの種類を表1に示す。カテゴリの種類の中から「その他の料理」を抜いた21種類を使用する。これはぐるなびの大分類である。

これをカテゴリ21種類それぞれに8人~10人を被験者として対象のカテゴリに対して，似ているカテゴリを上位5件を選択，似ていないカテゴリを下位5件を選択してもらう。この結果から対象のカテゴリに対して被験者達がつけた他のカテゴリの順位を値として上位5件(1~5)と下位5件(16~20)の値

表1 カテゴリデータ

C_1	イタリアン・フレンチ
C_2	居酒屋
C_3	焼肉・ホルモン
C_4	焼き鳥・肉料理・串料理
C_5	ダイニングバー・バー・ビアホール
C_6	洋食
C_7	日本料理・郷土料理
C_8	すし・魚料理・シーフード
C_9	お酒
C_{10}	和食
C_{11}	宴会・カラオケ・エンターテイメント
C_{12}	中華
C_{13}	アジア・エスニック料理
C_{14}	カフェ・スイーツ
C_{15}	鍋
C_{16}	お好み焼き・粉物
C_{17}	オーガニック・創作料理
C_{18}	欧米・各国料理
C_{19}	ラーメン・麺料理
C_{20}	ファミレス・ファーストフード
C_{21}	カレー
C_{22}	その他の料理

r を合計した値を算出した。しかし上位にも下位にも含まれない，つまり6位から15位までの r の値を10として今回は値を求めた。求めた合計の上位10件までの値に0.9から0.1までの0.1間隔で重みの基準 I を決定し，実際につけられた合計の値が重みの基準の値に近いのかを以下の式で求め，対象のカテゴリ C_a との類似度 $S_c(C_a, C_b)$ を決定した。

$$I = \begin{cases} 1 - 0.1 \times R(\sum_{r \in R_p} r) & (R(\sum_{r \in R_p} r) \leq 10) \\ 0 & (R(\sum_{r \in R_p} r) > 10) \end{cases} \quad (3)$$

$$M = 1 - 0.1 \times \frac{\sum_{r \in R_p} r}{P} \quad (4)$$

$$S_c(C_a, C_b) = I \times (1 - |I - M|) \quad (5)$$

P はそのカテゴリ組を評価した被験者数であり， R_p はそのカテゴリ1組に対して被験者がつけた順位の集合である。 R はカテゴリ C_b に対し被験者がつけた r の合計が低い順の順位を返す関数である。

実験の結果から類似度の指標を作成し選択された店舗のカテゴリの組み合わせで類似度の最も高い値をカテゴリの類似度として求めた(表2)。例として和食のカテゴリがついた飲食店 X_c と日本料理・郷土料理のカテゴリがついた飲食店 Y_c との類似度は $S_c(X_c, Y_c) = 0.9$ となる。

4.1.2 平均価格の類似度

飲食店それぞれの平均価格 X_p, Y_p を使って類似度を表す。価格で差が少なければ類似度の値は大きくなる。また本研究では同じ価格の差でも金額が低い飲食店同士の差は類似度に大きく影響し，金額が高い飲食店同士の差は類似度あまり影響しない

表2 カテゴリーの類似度

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}	C_{13}	C_{14}	C_{15}	C_{16}	C_{17}	C_{18}	C_{19}	C_{20}	C_{21}	
C_1	1.00																					
C_2	0	1.00																				
C_3	0	0.76	1.00																			
C_4	0.14	0.71	0.81	1.00																		
C_5	0.54	0.72	0	0.31	1.00																	
C_6	0.89	0	0.22	0.42	0.40	1.00																
C_7	0	0.43	0.09	0.34	0	0	1.00															
C_8	0	0.44	0	0	0	0	0.85	1.00														
C_9	0.23	0.85	0.32	0.09	0.82	0.23	0.37	0.34	1.00													
C_{10}	0	0.58	0.30	0.40	0	0	0.90	0.76	0.41	1.00												
C_{11}	0.08	0.68	0.37	0.45	0.70	0	0	0	0.53	0	1.00											
C_{12}	0	0	0.34	0.50	0.45	0.26	0.08	0.08	0.23	0	0.29	1.00										
C_{13}	0	0.40	0.42	0.25	0.23	0	0	0	0	0	0.16	0.65	1.00									
C_{14}	0.73	0	0	0	0.28	0.62	0	0	0	0.10	0.26	0	0	1.00								
C_{15}	0	0.70	0.34	0.25	0	0	0.71	0.34	0.31	0.65	0.24	0.08	0.30	0	1.00							
C_{16}	0	0.57	0.61	0.47	0	0	0.56	0.16	0	0.41	0	0.08	0	0.15	0	1.00						
C_{17}	0.36	0	0.17	0	0.31	0.32	0.63	0.28	0	0.46	0.08	0.54	0.08	0.58	0.08	0	1.00					
C_{18}	0.87	0	0	0.38	0.24	0.75	0	0.23	0.08	0	0	0.52	0.43	0.29	0	0	0.45	1.00				
C_{19}	0	0.35	0	0	0	0.16	0.27	0	0	0.45	0	0.72	0.90	0	0.39	0.22	0	0.08	1.00			
C_{20}	0.50	0	0	0	0.15	0.67	0	0.17	0	0.39	0.18	0	0.38	0.77	0	0.30	0	0.15	0.56	1.00		
C_{21}	0	0	0	0	0.29	0.49	0	0	0	0	0.17	0.86	0	0	0.09	0	0.42	0.49	0.23	0.48	1.00	

いように設定している。例として、500 円の飲食店と 2000 円の飲食店では類似度は低く、8000 円の飲食店と 9500 円の飲食店では、差が同じ 1500 円でも類似度は高くなる。平均価格の類似度は以下の式で求める。

$$S_p(X_p, Y_p) = \frac{\text{Min}(X_p, Y_p)}{\text{Max}(X_p, Y_p)} \quad (6)$$

この式に先ほどの例を当てはめると、500 円と 2000 円での類似度は 0.25 となり 8000 円と 9500 円での類似度は 0.84 となり同じ金額の差でも類似度に違いが出る。

4.1.3 説明文の類似度

本研究では、ユーザが選んだ飲食店の説明文に現れる単語はユーザの好みを表現していると考えた。説明文の類似度は、コサイン類似度を用いて求める。飲食店それぞれの TF-IDF による文書ベクトル X_d, Y_d を用いて、それぞれノルム $|X_d|, |Y_d|$ と内積 $X_d \cdot Y_d$ を求めることでコサイン類似度を求める。多くの飲食店ではコサイン類似度で求めた値が、低かったため求めた値を 3 乗根することで他の類似度とのバランスを調整した。

$$S_d(X_d, Y_d) = \sqrt[3]{\frac{X_d \cdot Y_d}{|X_d| |Y_d|}} \quad (7)$$

4.1.4 飲食店の構造類似度の算出

3つの要素の類似度を用いて店舗同士の類似度 $S(X, Y)$ を以下の式で求める。 α, β, γ はどの要素を重視するかを決める重みである。

$$S(X, Y) = \alpha \times S_c(X_c, Y_c) + \beta \times S_p(X_p, Y_p) + \gamma \times S_d(X_d, Y_d) \quad (8)$$

本稿では $\alpha=0.4, \gamma=0.2$ とする。 β は価格の類似度に対する重みであり、価格はカテゴリに依存している関係から値は 0.4 \times カテゴリの類似度とする。

$$\beta = 0.4 \times S_c(X_c, Y_c) \quad (9)$$

表3 飲食店の関係

店舗番号	カテゴリ	平均価格	説明文上位3件
1	居酒屋	3500	牛タン, 仙台, 1番
2	居酒屋	4000	牛タン, 仙台, たん
3	焼肉・ホルモン	4000	牛タン, おつまみ, 仙台
4	居酒屋	3500	コース, 伊達, 素揚げ
5	和食	3500	作り, ハイクラス, リッチ

以下に、重みの決定方針を表3を例に説明する。 x は店舗番号1と2の組み合わせで平均価格が違う組み合わせとする。 y は店舗番号1と3の組み合わせでカテゴリと平均価格が違う組み合わせとする。説明文は2つのキーワードが一致しているため、ここでは説明文は一致していると考えた。 z は店舗番号1と4の組み合わせで説明文が違う組み合わせとし、 w は店舗番号1と5の組み合わせでカテゴリと説明文が違っている組み合わせとする。

x と z の飲食店の類似度を比べると、カテゴリが一致しているため、平均価格が一致しているため z の方が類似度は高いと考えた。また、 y と w の類似度を比べると説明文が一致している y 方が類似度が近いと考えた。このようにすべての組み合わせでの類似度の大小関係を求めた結果、飲食店の類似度関係を $z < x < y < w$ と仮定した。その関係性となるように、 α, β, γ を設定した。これら3つの要素がすべて一致していた場合、類似度 $S(X, Y)$ は最大の1.0となる。

4.2 グループが好む飲食店ペアの抽出

グループの嗜好に合致する店舗ペアを決定するために、構造類似度を用いて、嗜好スコアを求める。嗜好スコアは以下の式で求める。

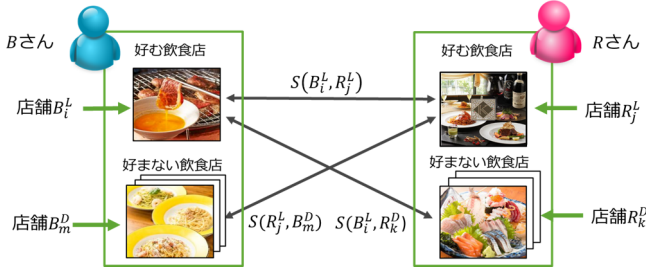


図 2 嗜好スコア の概念

$$\begin{aligned}
 \text{Score}(B_i^L, R_j^L) &= \frac{1}{4} \times (2 \times S(B_i^L, R_j^L) \\
 &+ (1 - \max_{R_k^D \in R^D} S(B_i^L, R_k^D)) \\
 &+ (1 - \max_{B_m^D \in B^D} S(R_j^L, B_m^D))) \quad (10)
 \end{aligned}$$

ここで、 B_i^L はユーザ 1 の好む店舗、 R_j^L はユーザ 2 の好む店舗、 B_m^D はユーザ 1 の嫌う店舗、 R_k^D はユーザ 2 の嫌う店舗である。 B^D, R^D はそれぞれのユーザの嫌う店舗の集合である。

例として 2 人のユーザを用いたものを (図 2) に示す。それぞれユーザ 1 の飲食店リストから飲食店 B_i^L と、ユーザ 2 の飲食店リストの飲食店 R_j^L の類似度 $S(B_i^L, R_j^L)$ を求める。次にユーザ同士の嫌いな飲食店を含めた類似度をそれぞれ求めるユーザ 1 が嫌いな飲食店リストの中で飲食店 R_j^L と類似度を求め一番類似度が高い飲食店を B_m^D とし、嫌いな飲食店を含めた類似度は $S(R_j^L, B_m^D)$ と表す。次にユーザ 2 が嫌いな飲食店リストの中で飲食店 B_i^L と類似度を求め一番類似度が高い飲食店を R_k^D とし、嫌いな飲食店を含めた類似度は $S(B_i^L, R_k^D)$ と表す。

4.3 グループに適した飲食店の推薦

グループに適した飲食店を推薦する流れを示す。

- (1) 嗜好スコアの最も高いペアを各ユーザペアの総当たりに対して抽出する。
- (2) ペアに含まれる店舗集合中に過半数となるカテゴリがある場合は 3 へ、そうでない場合は 4 へ。
- (3) ペアのカテゴリが一致していた場合その一致したカテゴリと、どちらかの平均価格を推薦条件とする。
- (4) 一致していない場合はランダムに選ばれた店舗のカテゴリとランダムな平均価格とする。
- (5) 3 か 4 の条件で、複数の飲食店を抽出する。抽出された飲食店の中で両方の飲食店との説明文の類似度を求める。説明文の類似度が最も高い飲食店を、グループに適した飲食店として推薦する。

5. 構造類似度の評価実験

5.1 実験データ

本研究ではぐるなび API を用いて飲食店データ 561104 件を取得し、その中でも新宿エリア 4276 件のデータを利用する。

使用するカテゴリの種類は表 1 にある、22 種類を使用する。1 つの飲食店のデータにはカテゴリは最大 2 つまで付与されて

表 4 実験で用いる店舗リスト

s ₁	天井てんや 新宿センタービル店
s ₂	もて茄子や 新宿御苑前
s ₃	リゾットカフェ 東京基地 ルミネエスト新宿店
s ₄	S A L A S
s ₅	お好み焼 本陣
s ₆	グリルレストラン&スイーツカフェ スコール 東新宿店
s ₇	ハクリュウカン wine & curry
s ₈	九州 熱中屋 西新宿 LIVE
s ₉	松尾ジンギスカン 新宿三丁目店
s ₁₀	太陽のトマト麺 新宿東宝ビル店

表 5 店舗ごとの相関係数の平均値

	平均類似度	カテゴリ類似度	提案手法
s ₁	0.66	0.61	0.66
s ₂	0.35	0.49	0.38
s ₃	0.47	0.48	0.44
s ₄	0.47	0.47	0.55
s ₅	0.42	0.43	0.42
s ₆	0.28	0.16	0.28
s ₇	0.29	0.14	0.23
s ₈	0.59	0.62	0.74
s ₉	0.64	0.66	0.62
s ₁₀	0.15	0.06	0.06
平均	0.43	0.41	0.44

いる。

本研究ではぐるなび API にある飲食店の説明文を用いる。説明文には 2 種類あり、短い説明文と長い説明文を用いて飲食店の説明文を取り出す。特徴語を取り出すために、飲食店の紹介文をまず MeCab [10] を用いて形態素解析を行い名詞を抽出する。このとき不要な語として数値と記号は取り除く。またカテゴリ名は飲食店の紹介文に含まれるが、すでにカテゴリという別属性として扱っているため、不要と判断し、カテゴリ名を省く。

5.2 実験内容

4.1.4 節で述べたシステムの店舗間の類似度に対して精度を求める実験を行った。本稿では、被験者に店舗の URL から情報を取り入れ、対象の店舗に対し他の店舗 10 店舗を似ている順番に 1 位から 10 位まで並べ替えてもらった。システムが対象の店舗に対して並べ替えた結果と被験者が並び替えた結果を以下のスピアマンの順位相関係数で評価した。

$$\text{Spearman} = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (x_i - y_i)^2 \quad (11)$$

今回は比較手法に各類似度の平均 ($\alpha = \beta = \gamma = 0.33$) による手法 (今後平均類似度と記載) とカテゴリの類似度のみ ($\alpha = 1.0$) による手法を用いた。

対象の店舗は表 4 の 10 店舗を用いた。これらの店舗は全てカテゴリが異なっている。

5.3 実験結果と考察

実験の結果を表 5 に示す。各値は対象の店舗とその全ての店

舗のに対する相関係数の平均を表している。各列は平均類似度との順位相関係数、カテゴリ類似度との順位相関係数、構造類似度を使って並べ替えた結果（今後提案手法と記載）との順位相関係数となっている。

全体を通して、表5の全ての結果の平均を見ると提案手法が0.44となり他の2手法をわずかに上回る結果となった。以下詳細を述べる。

順位相関係数の平均が一番高かった結果 s_8 の「九州 熱中屋 西新宿 LIVE」について結果と考察を述べる。結果として、提案手法で並べ替えた結果0.74と高い相関があることが得られた。平均類似度の場合0.59となり、カテゴリの類似度では0.62となり提案手法が大きく上回る結果となった。考察として、構造類似度を考慮した結果「九州 熱中屋 西新宿 LIVE」は被験者との並べ替えの結果に8割以上が0.7を超える値となったため相関が高くなった。カテゴリには「中華」と「鍋」が付与していて、並べ替えの1位の、「かば屋 歌舞伎町輝ビル店」ではカテゴリに「居酒屋」と異なっているためカテゴリの類似度は低かった。しかし店の説明文が、「九州 熱中屋 西新宿 LIVE」では居酒屋の説明文にもあるような単語（個室、飲み放題）などが多く含まれていたため多くの被験者と考えが一致したことから相関が高かったと言える。このことからカテゴリのみや平均類似度の並べ替えより構造類似度が高くなったとされる。

順位相関係数の平均が一番低かった結果 s_{10} の「太陽のトマト麺 新宿東宝ビル店」について結果と考察を述べる。結果として、提案手法で並べ替えた結果0.06とほぼ相関がない結果となった。平均類似度の場合0.15となり、カテゴリの類似度では0.06となった。考察として、相関係数の結果が低くなった原因は、全ての被験者で並べ替えとの結果に大きく差があることから低くなったと考える。並べ替えの結果と大きく異なった理由として「太陽のトマト麺 新宿東宝ビル店」ではカテゴリに「ラーメン・麺料理」と「オーガニック・創作料理」が付与されている。しかし並べ替える店舗に「バーミヤン 新宿西口大ガード店」がありその店舗は提案手法が1位と示し、被験者の6割は1位としているが、それ以外の店舗が対象の店舗と並び替えるのが人によってばらばらになり統一感がなかったため低い相関関係となった。被験者では並び替える判断が難しかったと考える。

6. グループが好むペアに基づく推薦結果と考察

4.2節で述べた、グループが好むペアを求める方法の有効性を考察するためにいくつかの例を示す。本稿では、2人のユーザの好む飲食店を10件ずつ仮想的に決定し、好まない（嫌いな）飲食店を2件ずつ仮想的に決定した（表6）。今回は各ユーザの飲食店選択を2つのパターンから出力の違いを比較する。2つのパターンから類似度の高かった上位3件の組み合わせとスコアを示す。

6.1 カテゴリが一致している飲食店が含まれている場合

カテゴリが一致している飲食店同士の類似度を求める。そのためユーザの好む飲食店のカテゴリを統一し、好まない飲食店はランダムで行った（表7）。システムが推薦した飲食店は、

表6 選択された飲食店の例

	ユーザ1が選んだ飲食店	ユーザ2が選んだ飲食店
ユーザが好む飲食店	京都おぼんざい 茶茶花 新宿	海鮮居酒屋 はなの舞 新宿西口パレット店
	バンビ 新宿サブナード店	炭火焼だいにんぐ「わたみん家」 新宿歌舞伎町店
	楽蔵 新宿東口店	とんかつ 和幸 新宿サブナード店
	人形町今半 新宿小田急第一生命ビル店	タイ料理 ティー・スズキキッチン 西新宿店
	和庵	讃岐うどん おごっと新宿南口店
	正月屋 吉兆	新宿 創作和食 Dining 御楽・GORAKU-
	レタしゃぶ 郷土料理 吟	スリーバー
	かがり	がぶ呑み居酒屋 薩摩八郎 新宿店
	パバミラノ 新宿三井ビル店	とり割烹 博多 華味鳥 新宿マルイアネックス店
	鍋ぞう 新宿三丁目店	個室肉バル KAGA・加賀・新宿東口本店
ユーザが好まない飲食店	ル・プラン 新宿店	寿司 高瀬
	代々木 ビザ 貸切 ROFL	焼鳥屋 鳥貴族 新宿3丁目店

表7 同じカテゴリによる嗜好スコア

店舗名	店舗名	店舗間の類似度	嗜好スコア
Tokyo Cafe 202	トンカツの店 豚珍館	0.75	0.69
ラムしゃぶ 金の目 新宿店	きづなすし 西新宿店	0.79	0.65
Tokyo Cafe 202	なか卯 新宿職安通り店	0.68	0.64

表8 同じカテゴリで飲食店の推薦

店舗名	カテゴリ	平均価格	一致している説明文
Tokyo Cafe 202	宴会・カラオケ・エンターテイメント、和食	1000	時
トンカツの店 豚珍館	和食	800	時、ボリューム
推薦された飲食店			
鉄板牧場 新宿店	和食	1000	ボリューム

「NIGORIZAKEBAR 濁酒本舗 Tejimaul」となった。

考察を以下に述べる。カテゴリが一致している飲食店同士の場合、表7の結果から、カテゴリが一致しているため比較的に類似度も高くなっていることがわかる。このことからユーザ同士が同じカテゴリの飲食店を好んでいく場合には推薦されやすい結果になる。上位1位の「Tokyo Cafe 202」と「トンカツの店 豚珍館」の組み合わせは説明文による一致はなかったが、カテゴリが一致し平均価格がほかの飲食店と比べて、差がなかったため類似度が一番高くなったと考える（表8）。

カテゴリが一致している場合での推薦結果「鉄板牧場 新宿店」は、表8の結果から説明文ではボリュームが一致しカテゴリ、価格が一緒の飲食店が推薦された結果となった。

6.2 異なるカテゴリで偏らせた飲食店選択の場合

異なるカテゴリでの飲食店同士の類似度を求める。今回はカテゴリが「カフェ・スイーツ」に偏っているユーザ、カテゴリが「ラーメン・麺料理」に偏っているユーザと設定して行った（表9）。システムが推薦した飲食店は「CafeMarNero (NY.)」となった。

考察として、異なるカテゴリでの飲食店同士の類似度に着目する。表9の出力結果から、カテゴリが一致していないため、類似度も低くなっている事がわかる。0に近い値になっている組み合わせでは、説明文による類似度の一致があることがわかる。また飲食店同士の類似度が0に近い状態になっても嗜好スコアの値が0.1よりも大きい値になる結果も得られた。主な理由としては、好まない飲食店との類似度が低い店舗は、数値が引かれないため値が大きくなっている。上位1位の「ハウスティ アルファカーム」と「麓屋 京王プラザホテル」では異なるカテゴリの中でも説明文の「等」が一致していたことから値が高くなっていると考える（表10）。

表 9 異なるカテゴリによる嗜好スコア

店舗名	店舗名	店舗間の類似度	嗜好スコア
ハウスティ アルファカーメル	麓屋 京王プラザホテル	0.10	0.47
PLATINUM BUFFET 小田急新宿ミロード	九龍餃子房 新宿別館	0.15	0.46
SCOPP CAFE	麓屋 京王プラザホテル	0.11	0.46

表 10 異なるカテゴリでの飲食店の推薦

店舗名	カテゴリ	平均価格	一致している説明文
ハウスティ アルファカーメル	カフェ・スイーツ	1000	等
麓屋 京王プラザホテル	和食, ラーメン・麺料理	5000	ランチメニュー, ランチ, 毎日, 空間, コース, 等
推薦された飲食店			
CafeMarNero (NY.)	カフェ・スイーツ, ダイニングバー・バー・ビアホール	5000	ランチ, 毎日, 空間, ランチメニュー, コース

異なるカテゴリでの推薦結果「CafeMarNero (NY.)」では表 10 の結果から, カテゴリ, 平均価格は「カフェ・スイーツ」「5000」と選ばれたが説明文は「空間, ランチメニュー, コース」などが一致し, カテゴリや価格も一致していることから推薦されたと考える。

7. まとめと今後の課題

本研究ではグループで個々が飲食店を選択し, その構造類似性に基づく飲食店推薦システムを提案した。グループでどんな飲食店を選ぶか悩んでいる際に, 好みや予算などをまとめるために飲食店同士での類似度を表す手法を提案した。飲食店の類似度を決める際に, お互いが好まない飲食店も考慮することでより推薦の精度を向上させ, グループに適した飲食店を求めた。

構造類似度の実験では提案手法が他の 2 手法をわずかに上回る結果となった。グループが好むペアの有効性実験では, 2 つのパターンでグループに適した飲食店を求めた。その結果から同じカテゴリが含まれている場合は比較的飲食店同士の類似度は高くなるが, 好まない飲食店を計算に加えることで, 出力される値に違いを出すことができた。またカテゴリに偏りがあるユーザ同士でも説明文を用いることで飲食店の類似度が異なるため, こちらも値の違いが見られた。

今後の課題として 3 つの要素以外にも写真などほかの要素で特徴をつかみ構造類似度の精度向上をはかる。今回はカテゴリの類似度のみを 1 にするという比較手法を用いたが, 他の要素の重みのみを 1 にして行い要素間の特徴をとらえ精度の向上をはかる。

また実際にユーザの履歴を用いて, 好みの判定を行い推薦された店舗がグループに適しているかを検討する必要がある。

謝 辞

本研究の一部は, 平成 28 年度科研費若手研究 (B)(課題番号: 15K16091) によるものです。ここに記して謝意を表すものとします。

- [1] 日本政策金融公庫国民生活事業本部生活衛生融資部: 外食に関する消費者意識と飲食店の経営実態調査. https://www.jfc.go.jp/n/findings/pdf/seikatsu25_1218a.pdf, 2013.
- [2] 瀬古俊一, 八木貴史, 茂木学, 武藤伸洋, 小林透. 行動履歴と嗜好に基づくグループ向けコンテンツ推薦手法の提案. 情報処理学会論文誌コンシューマ・デバイス&システム (CDS), Vol. 2, No. 3, pp. 56-65, 2012.
- [3] 伊藤冬子, 廣安知之, 三木光範. Ec サイトにおけるユーザの嗜好の変化の検出. 人工知能学会 論文集, Vol. 23, pp. 1-4, 2009.
- [4] 奥蘭基, 将史牟田, 廣美平野, 宗益子, 准一星野. 複数人での旅行における嗜好分析による観光地推薦システムの提案. 研究報告ヒューマンコンピュータインタラクション (HCI), Vol. 2015, No. 19, pp. 1-8, 2015.
- [5] 神畷敏弘. 推薦システムのアルゴリズム (2). 人工知能学会誌, Vol. 23, No. 1, pp. 89-103, 2008.
- [6] 大坪五郎. Gads-変化し続ける興味に対応する情報推薦. WISS2005 論文集, pp. 309-319, 2005.
- [7] 山田敬三, 西野香澄, 高木正則, 佐々木淳. 多人数のための飲食店決定支援システムにおける嗜好推定手法の提案-対話型進化計算に基づく主張区間推定法の提案-. In *DEIM Forum 2016*, No. H3-1, 2016.
- [8] 高玉圭樹, 佐藤史盟, 大谷雅之, 服部聖彦, 佐藤寛之, 山口智浩. 別カテゴリ商品提示による好みの明確化を促す推薦システム. 人工知能学会論文誌, Vol. 28, No. 2, pp. 210-219, 2013.
- [9] 山本真史, 山崎俊彦相澤清晴. Bag of words と skip-gram 併用によるレビュー・店舗間類似度評価とそれに基づく店舗推薦. 情報処理学会第 78 回全国大会論文集, pp. 543-544, 2005.
- [10] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. Applying conditional random fields to japanese morphological analysis. In *Proceedings of EMNLP 2004*.